Visual Computing Lab

Visualization of the Most Effective Features Extracted by ConvNets for Detection Tasks

Maximilian Kircher, Advisor: Faraz Saeedan

TU Darmstadt, Visual Inference

In this lab, the task was, to apply the visualisation methods presented in [1] to the object detection network of [2].

1 Class Model Visualization

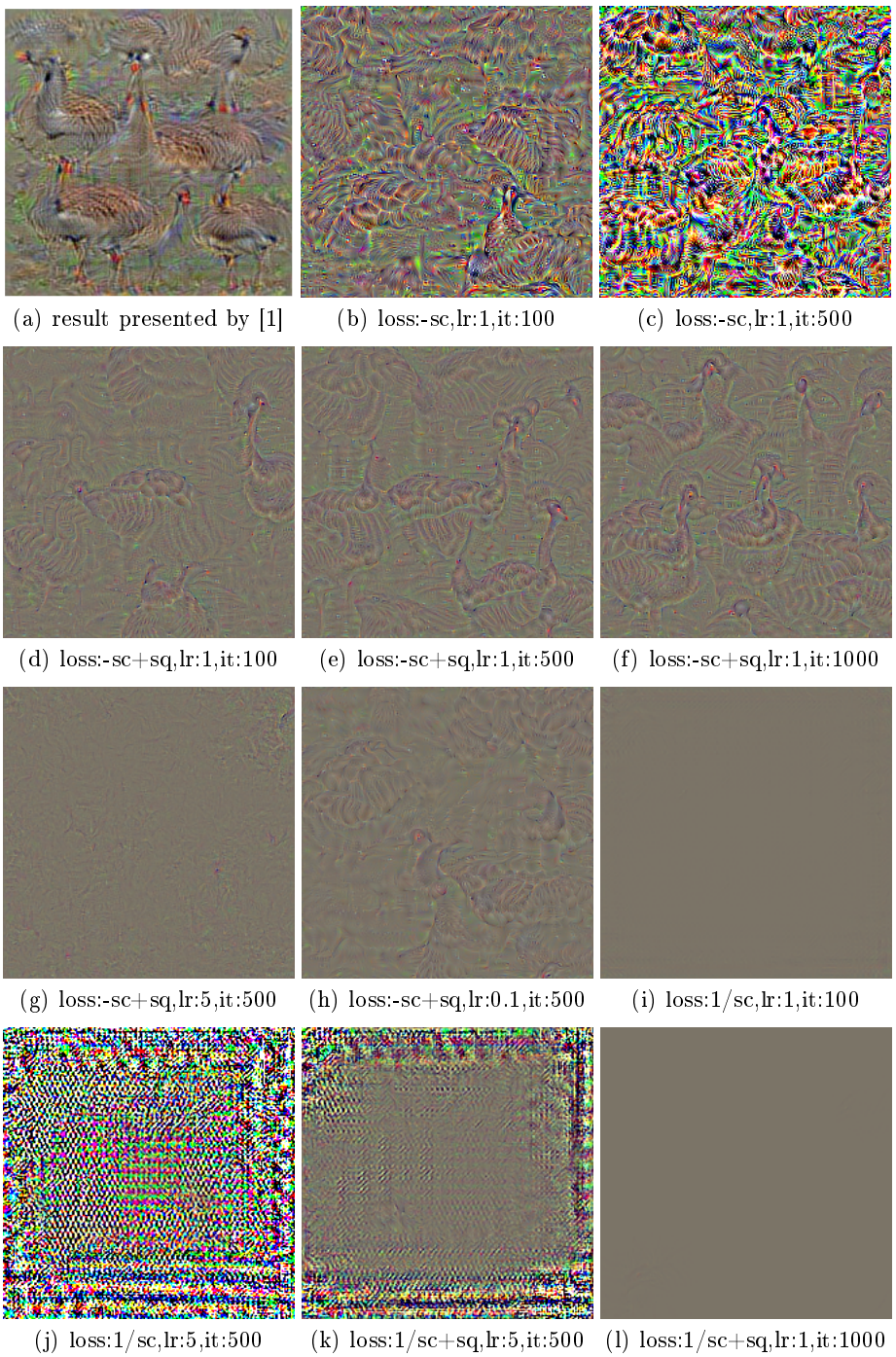
The first method of the paper is the class model visualization. Here the structures, that lead the network to classify an image in a specific class shall be visualized. With a classical backward pass with respect to the input image, the gradients are computed, that can then be used to optimise the image.

1.1 VGG

First I tried to reproduce the results from [1] with a vgg11 network as proposed in [?]. As examples I used the category *goose*, as an example of this category is also provided by the paper. (1 a)
The objective of the optimization is, to maximize the score for this categorization network. One naïve loss function would therefore be $-score(category)$ (1 b). One problem is, that the output of the network - i.a. the score - seems to often increase, with the input. Therefore the it is possible, that this loss function leads to chaos (1 c). A solution is an additional loss, that can be added to the loss function and ensures, that the values of the image don't become to big: Each pixel is squared and the mean over all pixel values is computed. (short: sq) (1 d-h)
With the combined loss, the methods produces good results, that are comparable to the result of the paper.
It can be seen, that a learning rate of 1 produces good results, while the results with to high or low learning rates are not that good (1 g,h). Another loss function instead of the negative class score, that came to my mind, was $\frac{1}{score(category)}$. But experiments with this function showed, that it seems to be not useful for the task (1 i-l).

1.2 SSD

For this network I chose the category *horse*, as the implementation I used was trained for other categories, than the 1000 image net categories.



(a) result presented by [1] (b) loss:-sc,lr:1,it:100 (c) loss:-sc,lr:1,it:500

(d) loss:-sc+sq,lr:1,it:100 (e) loss:-sc+sq,lr:1,it:500 (f) loss:-sc+sq,lr:1,it:1000

(g) loss:-sc+sq,lr:5,it:500 (h) loss:-sc+sq,lr:0.1,it:500 (i) loss:1/sc,lr:1,it:100

(j) loss:1/sc,lr:5,it:500 (k) loss:1/sc+sq,lr:5,it:500 (l) loss:1/sc+sq,lr:1,it:1000

Figure 1. Results of the class model visualization a vgg network for the category goose. lr is the used learning rate, it the number of iterations

As the ssd network is an object detection network and no categorization network, the loss function, that I used for the vgg network, can not be used directly here. I tried different loss functions:

1. There are different so called prior boxes used in the network, that determine, where it looks for the objects. I tried to use the biggest of them and maximize the category score as for the vgg network and added an additional term, that should ensure, that the prediction for this box is localized at the whole image That produced quite good results, but it can be seen, that it still concentrates on the upper left part, where the prior box is. (see Fig 2 a-c)

2. The second loss function I tried was, to maximize the category score for all prior box prediction. In the result, can be seen, that there are very many little structures, that might be horses, but that are not that clear (see Fig 2 d). When the optimization is continued with a smaller learning rate, it becomes even less clear, but there arise some slightly bigger ones, that are also more clear. (see Fig 2 e,f)

3. Third I used the criterion, that was also used, when training the network. It computes an localization and an confidence loss for a given target, for which I chose the wanted category and as boundaries the whole image. This loss produced results similar to the first loss, but it can be seen, that the optimization is relative costly. (see Fig 2 g-i)

4. This loss I used, to visualize a smaller (1/4) object at the center of the image. There are no big differences to the bigger version... TODO

   TODO decrease lambda for l3?

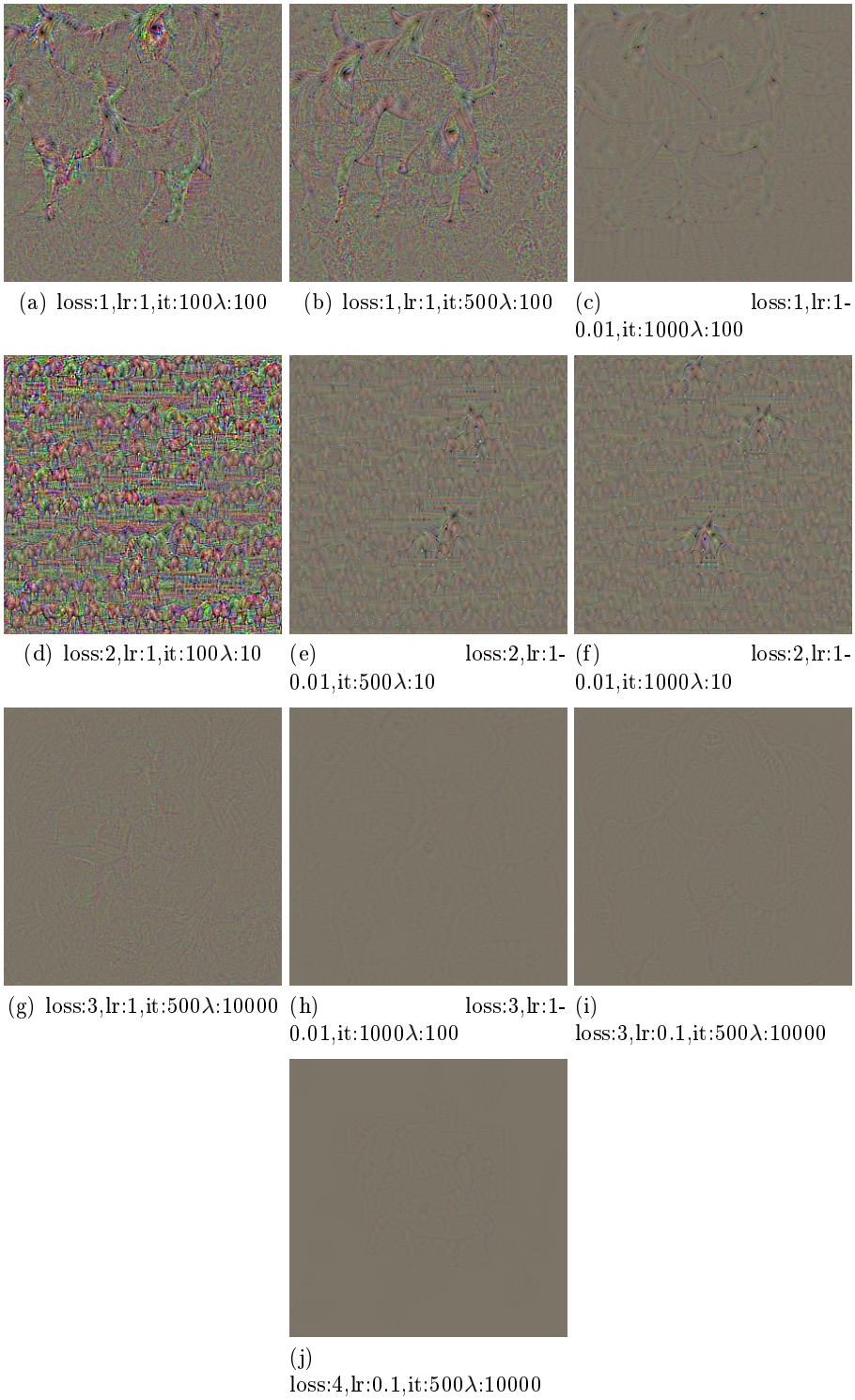
2 Image-Specific Class Saliency Visualisation

The second method presented in [1] is a saliency visualization: It is computed, which parts of a given image are important, to classify it to a given class. To do that, again the same gradients are computed, but this time, they are not used for optimisation. It is assumed, that pixels, that have high gradients are especially important for choosing the given class and therefore contain the object.

2.1 SSD

References

1. Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.
2. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.



(a) loss:1,lr:1,it:100λ:100 (b) loss:1,lr:1,it:500λ:100 (c) loss:1,lr:1-0.01,it:1000λ:100

(d) loss:2,lr:1,it:100λ:10 (e) loss:2,lr:1-0.01,it:500λ:10 (f) loss:2,lr:1-0.01,it:1000λ:10

(g) loss:3,lr:1,it:500λ:10000 (h) loss:3,lr:1-0.01,it:1000λ:100 (i) loss:3,lr:0.1,it:500λ:10000

(j) loss:4,lr:0.1,it:500λ:10000

Figure 2. Results of the class model visualization a vgg network for the category goose. lr is the used learning rate, it the number of iterations